## UNITED STATES DISTRICT COURT 1 EASTERN DISTRICT OF CALIFORNIA 2 SACRAMENTO DIVISION 3 4 Case No. 2:24-cv-02527-JAM-CKD CHRISTOPHER KOHLS, 5 **DECLARATION OF** Plaintiff, 6 SUPPORT OF PLAINTIFFS' MOTION FOR SUMMARY JUDGMENT AGAINST 7 v. **AB 2655** ROBERT A. BONTA, et al., 8 Judge: John A. Mendez Defendants. 9 10 Case No. 2:24-cv-02787-JAM-CKD THE BABYLON BEE, LLC, et al., 11 Plaintiffs, 12 v. 13 ROBERT A. BONTA, et al., 14 Defendants. 15 Case No. 2:24-cv-03315-JAM-CKD RUMBLE, INC., et al., 16 Plaintiffs, 17 v. 18 ROBERT A. BONTA, et al., 19 Defendants. 20 21 Case No. 2:24-cv-03162-JAM-CKD X CORP., 22 Plaintiff, 23 v. 24 ROBERT A. BONTA, et al., 25 Defendants. 26 27

28

DECLARATION OF

1	, being duly sworn, deposes and states as follows:
2	1. I am a at X Corp., covering product, policy, and
3	operations. I am submitting this declaration in support of Plaintiffs' Motion for Summary Judgment
4	Against AB 2655 in the above-captioned action. I have personal knowledge of the facts set forth
5	herein, unless otherwise noted. If called upon as a witness, I could and would competently testify
6	to those facts.
7	I. Applicability of AB 2655 and Section 230 of the Communications Decency Act
8	2. I have reviewed and am familiar with the statute in this case: California's AB 2655,
9	which is codified in law at Cal. Elec. Code §§ 20510–20520. AB 2655 applies to X Corp. and its
10	X platform.
11	3. X Corp. owns and operates the X platform, which is a "large online platform," as
12	defined by Section 20512(h), because it is a "public-facing internet website," "web application,"
13	"digital application," and "video sharing platform" that had at least 1,000,000 California users
14	during the preceding twelve months.
15	4. X also satisfies the definition of a "social media platform," as that term is defined in
16	Section 22675 of the California Business and Professions Code, because it is a public internet-
17	based service that has users in California and meets both of the following criteria:
18	(1)(A) A substantial function of the service or application is to connect users in order
19	to allow users to interact socially with each other within the service or application.
20	(B) A service or application that provides email or direct messaging services shall
21	not be considered to meet this criterion on the basis of that function alone.
22	(2) The service or application allows users to do all of the following:
23	(A) Construct a public or semipublic profile for purposes of signing into and using
24	the service or application.
25	(B) Populate a list of other users with whom an individual shares a social connection
26	within the system.
27	(C) Create or post content viewable by other users, including, but not limited to, on
28	

message boards, in chat rooms, or through a landing page or main feed that presents the user with content generated by other users.

5. The X platform is also an "interactive computer service," as defined by 47 U.S.C. § 230(f)(2), because it is a "system" that "provides or enables computer access by multiple users to a computer server."

## II. Burdens of AB 2655

- 6. Given that AB 2655's requirements hinge on vague phrases the meaning of which are subject to interpretation, complying with AB 2655 would be incredibly difficult in practice. For instance, whether or not content is true or false (i.e., depicts a candidate, elections official, or elected official¹ "doing or saying something" that they "did not do or say"), or would "falsely appear to a reasonable person to be an authentic record of the content depicted in the media," can be, and often is, subject to reasonable disagreement, particularly in the context of election-related content. The same is true as to whether particular election-related content: constitutes satire or parody; would be "reasonably likely to harm the reputation or electoral prospects of a candidate"; and would be "reasonably likely to falsely undermine confidence in the outcome of one or more election contests."
- 7. In many instances, it would be next to impossible for X Corp. to make the determinations required by AB 2655 within the time limits provided by the statute. For example, to determine whether content portrays a candidate for elective office "as doing or saying something that the candidate did not do or say," X Corp. would need to determine whether the person portrayed in the content was a "candidate for elective office" and, to determine if the statute applied, would

<sup>&</sup>lt;sup>1</sup> Because AB 2655 does not define "elected official," it is unclear how far the term reaches. Even if it only covered elected officials in California—and it is unclear whether the term is so limited—it would cover at least 120 individuals, and has the potential to cover many more, given that California has 57 counties and 482 cities, each presumably with elected officials. *See* Ex. 1 (*Elected Officials*, California State Assembly, available at https://clerk.assembly.ca.gov/legislative-information/elected-

officials#:~:text=The%20California%20State%20Assembly%20has,years%20in%20the%20State%20Legislature (last visited Mar. 7, 2025)) ("The California State Assembly has 80 Members" and the "California State Senate has 40 members[.]"); *id.* Ex. 2 (*Cities in California*, BallotPedia, available at https://ballotpedia.org/Cities\_in\_California (last visited Mar. 7, 2025)) (there are 57 counties and 482 cities, towns, and villages in California, according to a 2022 study from the U.S. Census Bureau). If the term applies to all officials in the United States, there could be several hundreds of thousands of "elected officials." or even more.

20

21

22

23

24

25

26

27

28

need to know the date of the election in which they were involved. X Corp. would also need to determine whether the candidate had actually said or done the thing that the candidate was portrayed as doing. That will be extremely challenging in many cases, particularly for candidates in local elections in smaller jurisdictions.<sup>2</sup>

- 8. It would also be difficult in many cases for X Corp. to determine whether the content in question was "digitally created or modified" or whether it was done so to make it "falsely appear to a reasonable person to be an authentic record of the content depicted in the media." Again, X Corp. may frequently not be able to readily determine what the "authentic record of the content depicted in the media" actually is, as X Corp. may not have access to those records, which may not be publicly available, or X Corp. may be unable to confirm whether the publicly available records are authentic or inauthentic.
- 9. Similarly, it would be difficult for X Corp. to determine whether the digital alterations of the audio or visual media would count as "minor modifications" that are exempt from the statute's prohibitions. The same goes for determining whether—as required under the statute content would be "reasonably likely" to "harm the reputation or electoral prospects of a candidate" or "falsely undermine confidence in the outcome of one or more election contests." These are inherently subjective judgment calls that will be incredibly difficult to make at mass-scale and within the statute's tight timeframes.
- 10. The statute also creates time pressure that is next to impossible to satisfy in most instances. For example, the statute requires covered platforms to respond to reports of "materially deceptive content" within 36 hours (and provide a description of what actions the platform will or will not take) and to take action as to such content within 72 hours. See §§ 20515(a), 20513(b). Given the volume of election-related content on X and the difficulties discussed above, it would likely be impossible for X Corp. to comply with those time limits as to every report.

<sup>&</sup>lt;sup>2</sup> For the 2026 election cycle, there will be at least 866 people qualifying as "candidate[s]" under Section 20512(c). There will be at least two candidates for each the offices of governor, lieutenant governor, attorney general, secretary o f state, treasurer, controller, insurance commissioner, and superintendent of public instruction (at least sixteen total); ei ght candidates for the board of equalization; 104 candidates for the federal House of Representatives; 40 candidates for state senate; 160 candidates for state assembly; 57 county clerks; and 482 city clerks.

- 11. X Corp. does not have, nor am I aware of, any technical tool that could effectively make these determinations on an automated basis at scale. To that end, to comply with AB 2655, X Corp. would need to hire additional employees to perform content-moderation functions. X Corp. would also need to train its content-moderation review team how to research and assess whether certain content falls within the requirements of AB 2655.
- 12. The enforcement mechanism for AB 2655 provides that covered platforms can be sued by candidates for elective office, elected officials, or elections officials, and by the California Attorney General, any district attorney in California, or city attorney in California. *See* §§ 20515(b), 20516. In all cases, the lawsuits may seek injunctive or other equitable relief against the covered platform to compel removal of certain "materially deceptive content," require labeling of such content, or to enforce the reporting process required by the statute. *Id.* The statute does not authorize lawsuits for putting back or unlabeling content that should not have been removed or labeled. In other words, the statute tilts the field: the platforms can only get sued if they fail to remove or label content that should have been removed or labeled—not if they mistakenly remove or label content that should not be removed or labeled.
- 13. Given the time pressures imposed by the statute (discussed above), and the fact that the enforcement regime set up under the statute provides that lawsuits can be filed only for *failing* to remove or label content that arguably should have been removed or labeled (i.e., censoring too little) and not for putting back and unlabeling content that should not have been removed or labeled (i.e., censoring too much), AB 2655 will provide enormous incentives for platforms like X to err on the side of removing or labeling content that presents a close call. Erring on the side of censorship will effectively provide the platforms with immunity from liability. Erring in the other direction would risk substantial liability.
- 14. Compliance with AB 2655 would be highly burdensome given the amount of content on the platform in general (roughly 170 billion posts per year) and the quantity of election-related content that X Corp. will have to decide whether to remove or label. For example, from November 5–6, 2024, there were 942 million posts on X globally, which was an all-time global

daily record. See Ex. 3 (Global Government Affairs (@GlobalAffairs), X (Nov. 14, 2024, 9:57

AM), available at https://x.com/globalaffairs/status/1857075415640117628?s=46 (last visited Mar.

7, 2025)). In addition, in the first six months of 2024, when elections were occurring in the EU and

India, X Corp. received 224,129,805 user reports related to content on the platform. It would be

untenable and very costly for X Corp. to have to litigate, on an expedited timeline, its content-

moderation decision-making as to each report of "materially deceptive content" under the statute,

and AB 2655 would thus heavily incentivize platforms like X to avoid these disputes and their

12

13

16 17

15

18

19 20

21 22

23

24 25

26

27 28

15. X already has policies that address AI-generated or manipulated media on the X platform that are false or misleading. Those policies are different from, and in certain respects conflict with, those set forth in AB 2655. In that sense, AB 2655 substitutes the government's preferred policies about how content should be moderated on X for X's policies.

16. Under X's "Authenticity" Policy, which covers "Synthetic and Manipulated Media," users "may not share inauthentic media, including, manipulated, or out-of-context media that may result in widespread confusion on public issues, impact public safety, or cause harm ('misleading media')." Ex. 4 (Authenticity, X, available at https://help.x.com/en/rules-andpolicies/manipulated-media (last visited March 7, 2025)) at 9. Moreover, in "situations where [X is] unable to reliably determine if [the content] is misleading media, [X] may not take action." Id. How X enforces its Authenticity Policy "depend[s] on the severity of the violation as well as any previous history of violations." Id. at 10. One such way in which X may "take action" to enforce its policy is to label posts containing misleading media, to help people understand the content's authenticity and to provide additional context. For example, from August 5, 2024 to February 3, 2025, X labeled 3,700 posts under its Authenticity Policy (as well as its predecessor version).

17. Under X's Authenticity Policy—which is publicly available to all users of the platform as well as to the public generally—X considers, e.g., the following in determining whether to remove and/or label content:

- a. Whether the media is "significantly and deceptively altered, manipulated, or fabricated in a way that fundamentally changes its meaning and can result in widespread confusion on public issues, impact public safety, or cause serious harm"; and
- b. Whether the media is "shared in a deceptive manner or out-of-context or with intent to deceive people about the nature or origin of the content and can result in widespread confusion on public issues, impact public safety, or cause serious harm[.]" *Id*.
- 18. X's Authenticity Policy also makes clear that "[s]haring manipulated or out-of-context media in non-deceptive ways" is "[a]llowed" under the policy. *Id.* at 11.
- 19. Application of these policies requires X to make judgment calls about, for example, what is misleading or what is satire. Under AB 2655, if the government disagrees with X's judgment calls, X can be sued and forced to adopt the government's views on these controversial questions.
- 20. In my experience, much election-related content that is reported as false is (i) often merely satire or parody and/or (ii) could reasonably be considered valuable political commentary. Although such content would not violate X's current content-moderation policies, X would be incentivized under AB 2655 to err on the side of removing the content to avoid costly litigation.
- 21. In addition to the Authenticity Policy described above, the X platform also has features that, in practice, tell users whether particular pieces of content are false and/or generated using artificial intelligence and are potentially misleading. For instance, there is a feature on X called Community Notes that allows certain users to provide additional context, in the form of a "note," for particular posts. The note becomes visible to the public if enough contributors from different points of view rate the note as helpful. And, in recognition of the fast-paced nature of social media, X has accelerated Community Notes, and now indicates "Lightning Notes" that start appearing on posts within an hour of being proposed, or within an hour of the post itself going live. Any user can become a Community Notes "contributor" so long as they (i) have no recent X rule

1

4 5

6

789

11 12

10

1415

13

1617

19

18

2021

22

232425

26

2728

violations, (ii) joined X at least six months ago, and (iii) provide a phone number. Contributors are given alias profiles that are not linked to their main X profile, and X Corp. does not write or edit these notes.

- 22. In my experience, Community Notes already provide a robust counterspeech method to inform X users whether a particular piece of content is false and/or generated using artificial intelligence and potentially misleading. The same is true of posts or replies that users leave on particular pieces of content, since those posts or replies often discuss whether the content is authentic. In fact, in 2023 and 2024, Community Notes were utilized in connection with roughly tens of thousands of posts suspected of being created or enhanced with artificial intelligence. Based on the data we have, we expect the number of such Community Notes to increase going forward.<sup>3</sup> Meta has recently announced that it will employ a similar system on Facebook, Instagram, and Threads. *See* Ex. 10 (*Community Notes: A New Way to Add Context to Posts*, Meta Transparency Center (Feb. 20, 2025), available at https://transparency.meta.com/features/community-notes (last visited Mar. 7, 2025).
- 23. Here are two examples where the Community Notes feature has helped inform X users of false or digitally altered content. First, on April 27, 2024, the X user @MyLordBebo posted a screenshot of what appeared to be a real New York Post article, titled "Congress to Vote on Bill That Would Criminalize Questioning the Events Surrounding 9/11." See Ex. 11 (Lord Bebo (@MyLordBebo), X (Apr. 27, 2024, 11:21 AM), available at https://x.com/MyLordBebo/status/1784241592422776863 (last visited Mar. 7, 2025)). Through the Community Notes feature, a description was added to the post stating that "No such article

<sup>&</sup>lt;sup>3</sup> See also Ex. 5 (Matthew R. Allen et al., Characteristics of X (Formerly Twitter) Community Notes Addressing COVID-19 Vaccine Misinformation, 331 J. Am. Med. Ass'n 1670 (2024)); Ex. 6 (Mary Whitfill Roeloffs, X's Community Notes Accurately Corrected Vaccine Misinformation 97% Of The Time Last Year, Study Says, Forbes (Apr. 24, 2024, 11:00 AM), https://www.forbes.com/sites/maryroeloffs/2024/04/24/xs-community-notes-accurately-corrected-vaccine-misinformation-97-of-the-time-last-year-study-says/ (last visited Mar. 7, 2025)); Ex. 7 (Thomas Renault et al., Collaboratively Adding Context to Social Media Posts Reduces the Sharing of False News (2024), available at https://ssrn.com/abstract=4800565 (last visited Mar. 7, 2025)); Ex. 8 (Yuwei Chuai et al., Community notes reduce the spread of misleading posts on X (2024), available at https://osf.io/preprints/osf/3a4fe (last visited Mar. 7, 2025)); Ex. 9 (Chiara Patricia Drolsbach et al., Community notes increase trust in fact-checking on social media, 3 PNAS Nexus 1 (2024)).

18

19

20 21

22 23

24 25

26

27

28

exists from the New York Post." Second, on April 27, 2024, the X user @brookebay21 posted a photo of a group of individuals carrying Confederate and Nazi flags, along with one individual carrying an Israeli flag. Through the Community Notes feature, a description was added to the post stating that "This image has been digitally altered to include an Israeli flag where none existed in the original[,]" and directing the user to the non-augmented photo. See Ex. 12 (Peter Symons#NHSLove (@brookebay21), X (Apr. 27, 2024, 9:24 AM), available at https://x.com/brookebay21/status/1784212026723639503 (last visited Mar. 7, 2025)).

- 24. In fact, there is nothing stopping the California government—in particular (i) AG Bonta (@AGRobBonta; joined in June 2018), (ii) Secretary Weber (@DrWeber4CA; joined in January 2014), (iii) Gavin Newsom (@CAgovernor; joined in May 2009), or (iv) their staffs (using those X accounts or other accounts that meet the contributor requirements)—from utilizing the Community Notes feature or from leaving posts or replies about particular pieces of content, to inform X users that particular content is "materially deceptive" and/or generated using artificial intelligence and potentially misleading. This method would not interfere with the contentmoderation rights of covered platforms the way that AB 2655 does currently.
- 25. Based on publicly available information, it is clear that other covered platforms, such as Meta, YouTube, TikTok, Snapchat, and Google Search also have their own policies designed to address false, misleading, and/or manipulated media.
  - a. See Ex. 13 (How to Identify AI Content on Meta Products, Meta, https://www.meta.com/help/artificial-intelligence/how-ai-generated-content-isidentified-and-labeled-on-meta/ (last visited Mar. 7, 2025)) ("Meta requires an AI label when content has photorealistic video or realistic-sounding audio that was digitally created, modified or altered, including with AI.");
  - b. See Ex. 14 (Disclosing Use of Altered or Synthetic Content, YouTube, https://support.google.com/youtube/answer/14328491 (last visited Mar. 7, 2025)) ("To help keep viewers informed about the content they're viewing, we require creators to disclose content that is meaningfully altered or synthetically

20

22

23

24 25

26

27

28

generated when it seems realistic.");

- c. See Ex. 15 (About AI-generated TikTok. Content. https://support.tiktok.com/en/using-tiktok/creating-videos/ai-generated-content (last visited Mar. 7, 2025)) ("We also require creators to label all AI-generated content that contains realistic images, audio, and video, as explained in our Community Guidelines.");
- d. See Ex. 16 (Generative AISnapchat, Snapchat, onhttps://help.snapchat.com/hc/en-us/articles/25494876770580-Generative-AIon-Snapchat (last visited Mar. 7, 2025)) ("We may indicate that a feature in Snapchat is powered by generative AI in a number of ways . . . When you see these contextual symbols or other indicators in Snapchat, you should know that you are . . . viewing content that has been produced using AI and does not depict real world scenarios.");
- e. See Ex. 17 (Google Search's guidance about AI-generated content, Google Search, https://developers.google.com/search/blog/2023/02/google-search-andai-content (last visited Mar. 7, 2025)) ("Using automation—including AI—to generate content with the primary purpose of manipulating ranking in search results is a violation of our spam policies.").
- 26. In addition to those policies, X and other covered platforms also have contentmoderation policies that specifically apply to election-related content. The requirements of those policies are all different in certain respects than the requirements set forth under AB 2655.
  - a. See Ex. 18 (Civic Integrity Policy, X, https://help.x.com/en/rules-andpolicies/election-integrity-policy (last visited Mar. 7, 2025)) ("You may not use X's services for the purpose of manipulating or interfering in elections or other civic processes, such as posting or sharing content that may suppress participation, mislead people about when, where, or how to participate in a civic process, or lead to offline violence during an election. . . . [But] [n]ot all false or

untrue information about politics or civic processes constitutes manipulation or interference . . . the following are generally not in violation of this policy: inaccurate statements about an elected or appointed official, candidate, or political party[.]");

- b. *See* Ex. 19 (*Misinformation*, Meta, https://transparency.meta.com/policies/community-standards/misinformation (last visited Mar. 7, 2025)) ("We remove misinformation where it is likely to directly contribute to the risk of imminent physical harm. We also remove content that is likely to directly contribute to interference with the functioning of political processes. In determining what constitutes misinformation in these categories, we partner with independent experts who possess knowledge and expertise to assess the truth of the content and whether it is likely to directly contribute to the risk of imminent harm.");
- c. See Ex. 20 (Elections Misinformation Policies, YouTube, https://support.google.com/youtube/answer/10835034 (last visited Mar. 7, 2025)) ("Certain types of misleading or deceptive content with serious risk of egregious harm are not allowed on YouTube. This includes certain types of misinformation that can cause real-world harm, like certain types of technically manipulated content, and content interfering with democratic processes," including "Voter suppression" and "Incitement to interfere with democratic processes[.]");
- d. See Ex. 21 (Protecting Election Integrity in 2024, TikTok, https://newsroom.tiktok.com/en-us/protecting-election-integrity-in-2024 (last visited Mar. 7, 2025)) ("We don't allow manipulated content that could be misleading, including [AI-generated content] of public figures if it depicts them endorsing a political view.");
- e. See Ex. 22 (Harmful False or Deceptive Practices, Snapchat,

https://values.snap.com/policy/policy-community-guidelines/harmful-false-deceptive-information (last visited Mar. 7, 2025)) ("[W]e enforce our policies to apply to the following types of threats to civic processes:" "Procedural interference," "Participation interference," "Fraudulent or unlawful participation," and "Delegitimization of civic processes[.]");

- f. See Ex. 23 (Political Content, Google Search, https://support.google.com/adspolicy/answer/6014595?hl=en (last visited Mar. 7, 2025)) ("Advertisers must disclose all election ads that contain synthetic or digitally altered content by selecting the checkbox in the 'Altered or synthetic content' section in their campaign settings.").
- 27. As one would expect, each covered platform has a slightly different approach to moderating this type of content. AB 2655, however, would require all covered platforms—not just X—to comply with the government's blanket views about how to moderate this type of content.
- 28. It is very like that, for future elections in California, X will continue to curate and host content that is potentially covered under AB 2655. Moreover, X will continue its practice of not removing such content unless it violates X's content-moderation policies.

## IV. The Controversial Nature of Content Moderation

29. Social media content moderation is an inherently controversial undertaking. How social media companies apply their content-moderation rules to particular posts is a sensitive and controversial exercise that is subject to rigorous debate and controversy. Put another way, deciding what content should appear on a social media platform is a question that engenders considerable debate among reasonable people about where to draw the correct proverbial line. This is particularly true of election-related content and is certainly true of the type of content that is arguably covered by AB 2655's definition of "materially deceptive content." X Corp. recognizes the special importance of its users' ability to engage in political speech and dialogue on the X platform. Reasonable minds could disagree about what an appropriate content-moderation policy would be. Reasonable minds could also disagree about what the right way is to apply any of the

many platform content-moderation policies to any post on any of the platforms' sites or apps.

- 30. At X Corp., we take very seriously our responsibility to make well-thought-out content-moderation decisions and to make sure those decisions are accessible and understandable to our users. But in making these difficult calls, we have learned that, no matter what we do, some portion of the public will likely take issue with the way we have made these difficult judgment calls. To that end, it is imperative that X Corp. be able to make those decisions *itself*, free from looming threats of costly, expedited litigation.
- 31. Government officials and/or candidates, who often have a stake in the outcomes of elections, are particularly poorly situated to make decisions about what constitutes "materially deceptive content" about elections in a dispassionate way. Those government officials and candidates could easily use AB 2655 as a tool to impose burdens on the political speech of those with whom they disagree. And they could easily exercise their discretion not to bring any enforcement actions based on content that supports those policies and candidates with whom they agree. AB 2655 would give these officials the ability to impose substantial pressure on platforms, like X, to censor the speech of their political opponents and their supporters, by arguing that the speech is "materially deceptive" and otherwise subject to the statute.
- 32. X Corp. would thus be substantially harmed if forced to comply with AB 2655 against its will. The law would force X Corp. to take positions about the content permitted on its platform with which it disagrees. It would also force X Corp. to become a tool for private citizens' retaliation against other citizens' political speech.
- 33. Users of the X platform (and other covered platforms) would also be substantially harmed by AB 2655. The law would limit those users' access to information about political candidates and elections to content that the State deems appropriate for their viewing.

I declare under penalty of perjury of the laws of the United States of America that the foregoing is true and correct.

Executed on this  $\overline{\mathcal{Z}}$  day of March, 2025 in

